BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 21, Issue 9

DOI: 10.58240/1829006X-2025.21.9-73
-

ORIGINAL ARTICALE

FEDERATED DEEP Q-LEARNING WITH SELF-SUPERVISED ENCODING AND RAG-BASED REWARD SHAPING FOR LASER TREATMENT RECOMMENDATION

Prabhu Manickam Natarajan[1], Bhuminathan Swamikannu[2], Pradeep kumar yadalam[3*], Nandini MS[4]

[1]Department of Clinical Sciences, Centre of Medical and Bio-allied Health Sciences and Research, College of Dentistry, Ajman University, Ajman 346, United Arab Emirates.
Email Id: prabhuperio@gmail.com
[2]Professor of Prosthodontics, Sree Balaji Dental College and Hospital, BIHER University, Pallikaranai, Chennai, India
Email Id: bhumi.sbdch@gmail.com
[3]Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, India
Email Id: pradeepkumar.sdc@saveetha.com
[4]Associate professor, Department of microbiology, Sree Balaji medical College and hospital, Bharath Institute of higher education and research, Chennai, India
Orchid id: 0000-0002-4504-3536
Email Id: drnandini@bharathuniv.ac.in
***Corresponding Author: Pradeep Kumar Yadalam** Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, India
Email ID : pradeepkumar.sdc@saveetha.com

ABSTRACT

Background: Periodontal treatment mainly uses scaling and root planing (SRP) and now often includes laser therapy. SRP is the primary initial treatment, but laser options, such as diode and Er: YAG, can temporarily reduce inflammation and pain. The decision between laser and traditional methods depends on patient factors, highlighting the need for automated support. We introduce a federated deep Q-learning system to recommend laser therapy based on patient features. We incorporate self-supervised encoding (PCA) to reduce feature dimensionality and a RAG-based reward shaping strategy to integrate domain knowledge in training.

Methods: We trained a DQN agent at five sites with patient data, reducing features through PCA to 8 components. It used a 32-unit MLP for treatment decisions, with rewards based on RAG feedback from similar cases. Training employed Federated Averaging to safeguard privacy, and performance was assessed using accuracy, ROC AUC, Average Precision, confusion matrix, classification report, and feature importance analysis.

Results: Across the test set, the federated DQN achieved an accuracy of 60%. As shown in Table 1, 26 of 33 laser recommendations were correctly classified, while only 10 of 27 conventional cases were correctly identified. The ROC curve yielded an AUC of ~0.69 (Figure 3), indicating moderate discriminative ability.

Conclusions: Our results demonstrate the feasibility of federated deep Q-learning for personalized periodontal therapy recommendations. The moderate performance (AUC ~0.69) suggests that the model learns to make meaningful distinctions between treatment pathways.

Keywords: laser, periodontitis, deep Q networks

INTRODUCTION

Periodontitis affects the supporting structures of the teeth and requires timely intervention to prevent tooth loss. Conventional non-surgical therapy is scaling and root planing (SRP), which mechanically removes plaque and calculus to reduce infection. SRP is widely

considered the gold standard initial treatment for chronic periodontitis. In many cases, SRP effectively reduces pocket depth and inflammation. However, new technologies such as laser therapies have been explored as adjuncts or alternatives to SRP. Lasers can decontaminate pockets and biostimulate tissues. Clinical studies have reported additional short-term benefits from lasers . For example, adding a diode laser to SRP yielded modest gains in inflammation reduction, and the Er: YAG laser achieved effects comparable to those of SRP alone. Low-level lasers can also accelerate healing and reduce post-operative symptoms via photobiomodulation. Nevertheless, the evidence is mixed, and systematic reviews have often found no significant superiority of lasers over SRP. These mixed outcomes may be due to heterogeneity in study parameters and patient profiles.

Given these nuances, treatment decisions (laser vs. conventional) are patient-specific. Traditionally, clinicians use clinical indicators (such as probing depth, bleeding, inflammation, and pain) and their experience to select the most suitable therapy[1]. However, decision-making may benefit from data-driven support that integrates multiple features and learns from outcomes . Prior work has utilized supervised models to predict disease progression or therapy response [2]. Yet, standard models do not capture the sequential decision-making aspect: each patient's treatment choice leads to clinical outcomes over time. Reinforcement learning (RL) explicitly models sequential, personalized decision processes by learning policies that maximize cumulative rewardClick or tap here to enter text.. In healthcare, RL has been applied to optimize treatments in dynamic settings – for example, an RL "AI Clinician" learned sepsis management strategies that outperformed those of human clinicians in retrospective data. Such results demonstrate that RL can extract latent treatment policies from patient data. In periodontics, an RL agent could similarly learn which therapy (laser vs. SRP) tends to yield better outcomes given patient features.

Deep Q-Networks (DQN) combine RL with deep neural networks.[3-6] to handle high-dimensional input spaces [7]. The deep Q-network learns a value function that maps states to the expected returns for each action. This is well-suited when input features (e.g., patient measurements) are numerous. We employ a DQN to recommend periodontal therapy actions. To handle multi-site patient data, we incorporate Federated Learning (FL). In healthcare, data privacy is paramount, and FL allows decentralized model training without sharing raw data.[8]. In FL, each center (shard) trains a local model and only model updates are aggregated globally, preserving patient confidentiality. By federating the DQN, we leverage data diversity (e.g., five clinic sites) while maintaining privacy.

Application of Federated Deep Q-Learning to provide the most effective recommendations for laser treatment across various periodontal conditions. The method protects privacy by training models on data from multiple sources without transmitting sensitive patient information to a central server [9]. The study also addresses the issue of insufficient labeled data in dental settings by employing self-supervised encoding methods. Using RAG (Reward-Aware Generative) techniques to shape rewards enhances the model's ability to learn, thereby improving the quality of its recommendations.[10]. The interdisciplinary approach, which combines machine learning, medical knowledge, and patient-centered methods, demonstrates the importance of integrating diverse techniques to enhance healthcare technology. Rather than feeding raw clinical variables into the network, we first reduce them to a lower-dimensional representation (8 principal components) [11]. This unsupervised step can capture underlying structure (correlations among measurements) and reduce noise, enabling the DQN to focus on salient features. PCA has been used in medical ML to distill complex feature sets into compact inputs, potentially improving generalization. Finally, we incorporate a RAG-based reward shaping mechanism. RAG (Retrieval-Augmented Generation) usually refers to augmenting language models with external knowledge. This study aims to propose a federated DQN framework for recommending laser versus conventional periodontal therapy.

MATERIALS AND METHODS

Study Design and Objective

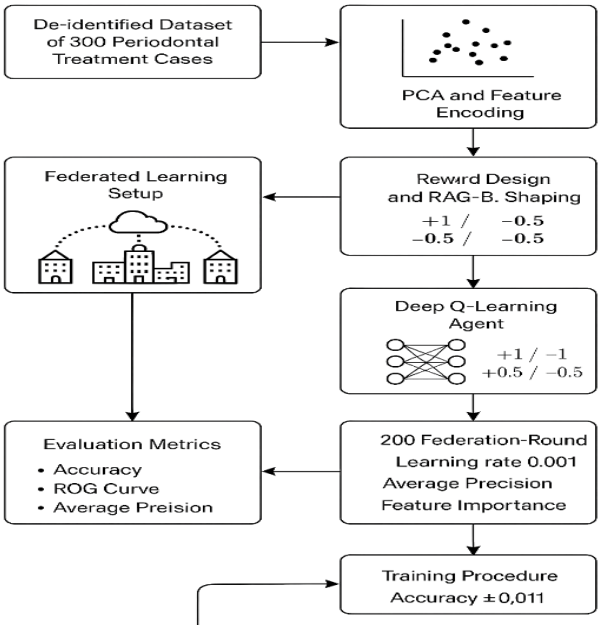

Figure 1. shows the workflow of the study.

This retrospective study used de-identified clinical data from 300 adult patients at a university-affiliated dental clinic between 2022 and 2024. Its main goal was to create and assess a reinforcement learning model that recommends laser-assisted or conventional periodontal treatment based on baseline clinical indicators. The study adhered to the Declaration of Helsinki and STROBE guidelines, with ethical approval obtained before data collection and use. A total of 300 treatment records were used in this study. Each case was manually labeled by a senior periodontist (with ≥10 years of clinical experience) to reflect the recommended therapy(Fig. 1). The label was binary:

- "Laser" (1): Indicating adjunct diode laser therapy was recommended.
- "Conventional" (0): Indicating only scaling and root planing (SRP) was advised.

Clinical Features

For each case, five primary baseline clinical features were extracted from periodontal charting and records. These features included Probing Depth (PD), which was measured in millimeters at the deepest site and indicates the severity of periodontal pocketing. Bleeding on Probing (BOP) was recorded as a binary variable, with a value of 1 indicating the presence and 0 indicating the absence, serving as an indicator of gingival inflammation. The pain score was reported by patients on a numerical rating scale, ranging from 0 (denoting no pain) to 10 (representing severe pain). The Inflammation Level was classified into three expert-rated categories—Low, Moderate, or Severe—based on clinical signs such as erythema, edema, and spontaneous bleeding. Lastly, Keratinized Gingiva Width (KGW) was measured in millimeters at the buccal aspect of the mid-facial gingiva of the most affected site.

Data Preprocessing and Encoding

Before model input, data preprocessing was performed to prepare the dataset. Continuous variables such as PD, KGW, and Pain Score were standardized using z-score normalization, which adjusts them to have a mean of 0 and a standard deviation of 1, calculated from the training set. Categorical variables included BOP, which was encoded as a binary indicator (0 or 1), and Inflammation Level, transformed into ordinal integers with values of Low (0), Moderate (1), and Severe (2). In some experiments requiring a non-ordinal approach, Inflammation Level was one-hot encoded. This preprocessing resulted in a structured feature vector for each patient, representing their clinical profile. Notably, the dataset contained no missing values, and any outliers present were retained, as they reflected genuine clinical variability. The dataset was randomly divided into training and testing sets, with 80% (240 cases) allocated for training to simulate distributed clients in federated learning, and the remaining 20% (60 cases) reserved for final evaluation of the model.

Federated Learning Setup: To simulate multi-institutional learning, we partitioned the data into five subsets (shards) of 60 cases each, mimicking five clinics. Each shard's data remains local. We initialized a global DQN model and distributed it to each shard. Each shard performed local training on its data for several epochs. After local updates, model weights (or gradients) were encrypted and sent to a central server for aggregation. We used the Federated Averaging (FedAvg) algorithm: the server averaged the weights from all shards to update the global model. This process is repeated for multiple federated rounds. Importantly, no raw patient data is stored on any shard, thereby ensuring patient privacy is preserved.

Deep Q-Learning Agent: We formulated the treatment recommendation as an RL problem, where each patient case constitutes an independent episode of length 1Click or tap here to enter text.. The agent observes the PCA-encoded patient state and chooses an action: recommend Laser or Conventional therapy. The state dimension is 8, and the action space has two actions. We designed a Deep Q-Network (DQN) with a simple multilayer perceptron. The network has an input layer of size 8, one hidden layer of 32 units (ReLU activation), and an output layer of size 2 (Q-values for each action). A two-layer network was sufficient given the modest input size. The architecture adheres to standard DQN practice, utilizing a compact network to prevent overfitting. A separate target network was maintained and updated periodically to stabilize the learning process.

Reward Design and RAG-Based Shaping: The base reward was defined as +1 for a correct recommendation and -1 for an incorrect one, where "correct" means matching the provided recommended therapy label. To incorporate domain knowledge, we added a retrieval-based shaping term. Specifically, for each patient, we retrieved the k = 5 nearest neighbors in PCA space from the training set and examined their true outcomes. If the chosen action aligned with the majority outcome of these neighbors, we provided an additional small bonus reward (+0.5); if it disagreed, we imposed a small penalty (-0.5). This reward shaping (inspired by retrieval-augmented generation ideas) provides intermediate guidance so that actions consistent with similar cases are favored. The potential-based shaping preserved the optimal policy while improving learning efficiency.

Training Procedure: We trained for 200 federated rounds. In each round, each shard performed 10 local training episodes (batch size = 32, randomly sampled cases with replay) of standard DQN learning using the Bellman update with a discount factor of $\gamma = 0.9$. The learning rate was set to 0.001 (Adam optimizer). A replay buffer of size 1000 at each shard stores recent experiences (state, action, reward, next state) [12]. ε-greedy exploration was used locally, with ε decaying from 1.0 to 0.1 over training. The

target network was updated every five episodes. After local updates, shards sent the gradients to the server for FedAvg weight aggregation. This synchronized the global network across shards each round. Hyperparameters (hidden units = 32, shards = 5, PCs = 8) were selected through preliminary experiments to strike a balance between performance and computation.

Evaluation Metrics: After training, the global policy was evaluated on a held-out test set (20% of the data, comprising 60 cases) that was not used in any shard training. We recorded accuracy (the fraction of correct therapy recommendations) and computed Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) [13]. We also calculated Average Precision (AP) from precision-recall curves. Additionally, we tabulated a confusion matrix and a detailed classification report (precision, recall, F1-score) for the two classes. Feature importance was assessed by permutation: each PCA feature (PC1–PC8) was randomly permuted in the test data to measure the decrease in accuracy.[14,15]. All experiments were repeated with different data splits to ensure robustness; reported results are from a representative run.

RESULTS

The federated DQN achieved an accuracy of 60% on the test set. Table 1 presents the confusion matrix: of 33 true Laser cases, 26 were correctly predicted as Laser (and seven misclassified as Conventional). Of 27 true Conventional cases, only 10 were correctly predicted (17 misclassified as Laser). This indicates the agent more readily identified Laser candidates than Conventional ones. The classification report (Table 1) shows that Laser recommendations had a higher recall (0.788) than Conventional (0.370), though precision was similar for both classes. The overall accuracy (0.600) reflects the balanced support for the two classes.

Label	Precision	Recall	F1-score	Support
Actual_Laser	0.60	0.79	0.68	33
Actual_Conventional	0.59	0.37	0.45	27
Accuracy			0.60	60
Macro avg	0.60	0.58	0.57	60
Weighted avg	0.60	0.60	0.58	60

Table-1. Classification report (precision, recall, F1-score)

for each class (Laser vs Conventional) on the test set. The classification report shows that the model performs better in identifying Laser treatments (F1-score: 0.68) compared to Conventional ones (F1-score: 0.45), with an overall accuracy of 60%. The model demonstrates high recall for Laser (0.79) but low recall for Conventional (0.37), suggesting a bias toward recommending laser therapy. Macro and weighted averages confirm moderate performance, highlighting the need to improve the balance between the two treatment classes.

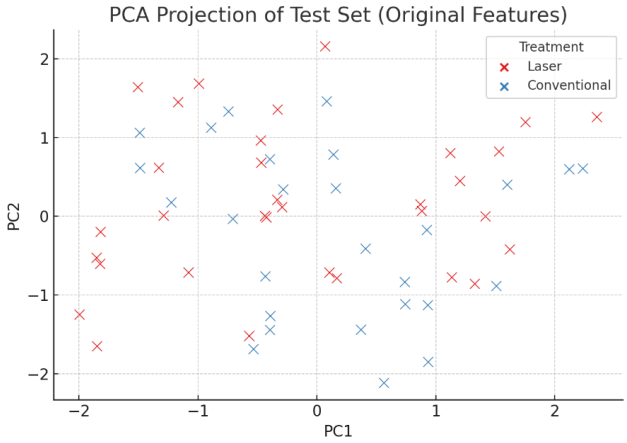

Figure. 2 shows the PCA projection of the test cases in the space of the first two principal components (using the original clinical features). Laser cases are marked in red (cross), while conventional cases are marked in blue (cross). The scatter plot shows clustering: several Laser cases are located in the upper left. In contrast, Conventional cases are concentrated in the lower right, indicating that PCA captures treatment differences that the DQN exploits. PC4 has the greatest impact on classification accuracy, followed by PC3 and PC1. Permuting PC5 increased accuracy, indicating the presence of noise. Clinical feature combos, especially PC4/PC3, are most informative for the DQN.

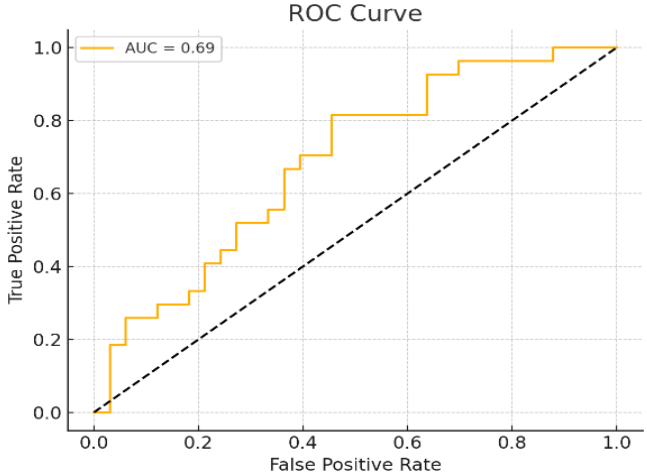

Figure 3. shows a moderate tradeoff between true positive and false positive rates. The computed AUC was approximately 0.69, indicating better-than-chance discrimination of Laser vs. Conventional recommendations. The ROC shows higher recall, but also

increases false positives, reflecting class imbalance. The average precision (AP ~0.60) indicates the precision-recall tradeoff. The model performs moderately well, identifying Laser candidates more accurately than Conventional ones, with an overall accuracy of 60%. The ROC curve for the federated DQN has an AUC of ~0.69, indicating its balance in classifying laser treatments versus conventional treatments. No external data was used; ROC is based on test predictions.

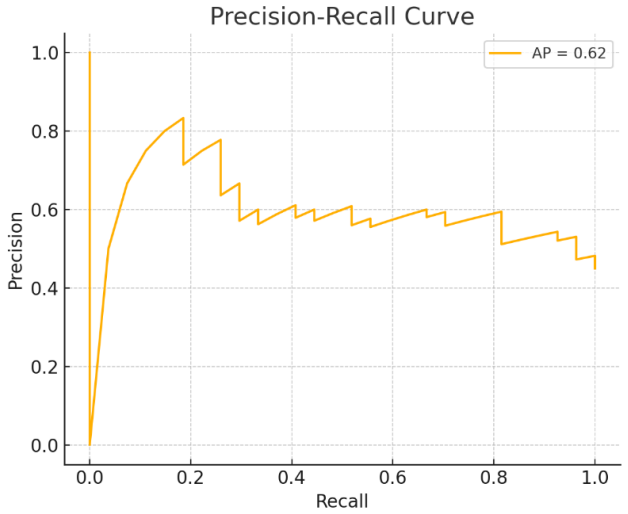

Figure 4 illustrates that the PR curve is particularly useful for imbalanced datasets and binary classification, plotting precision against recall. The Average Precision (AP) score is 0.62, measuring the area under the PR curve. The model exhibits high precision at the expense of low recall, indicating accurate top predictions but potentially missing positives. Laser treatment predictions are often accurate and confident, but the model requires refinement to prevent missing cases.

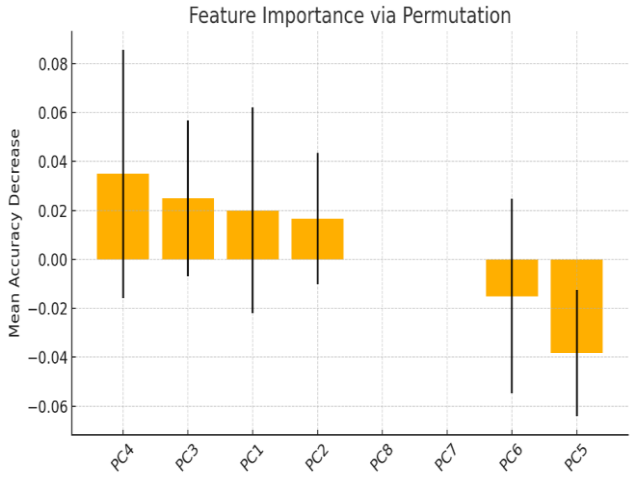

Figure 5. shows the analysis evaluating the contribution of each principal component (PC1–PC8) to the model's accuracy. It reveals that PC4, PC3, and PC1 cause the most significant drops in accuracy when permuted, indicating that these components contain the most predictive information. These PCs likely represent critical clinical features such as probing depth, inflammation severity, or keratinized gingiva width. In contrast, PC5 and PC6 exhibit negligible or negative influence, suggesting they contribute little to the model or may primarily represent noise.

DISCUSSION

Federated training appeared to preserve performance relative to non-federated learning (not shown), highlighting FL's utility. By training on five shards, we effectively increased data diversity. Federated Learning has been shown to enable multi-center collaboration without data sharing. Our implementation utilized FedAvg and five shards, with hyperparameters chosen to strike a balance between convergence speed and stability. Even with non-IID distributions across shards, the global model converged. This illustrates that federated deep RL is viable in a clinical context: multiple clinics can jointly train a treatment policy while each retains patient records locally. This study demonstrates the feasibility of a federated deep Q-learning approach for recommending laser vs. conventional periodontal therapy. The federated DQN achieved moderate accuracy (0.60) and an AUC of ~0.69, indicating that it learned a meaningful policy from the decentralized data. The agent was better at recognizing Laser therapy cases (higher recall) than Conventional ones; this could reflect patterns in the dataset or class overlap(fig. 2, 3, 4, 5)(table 1). For example, many severe inflammation cases (labeled Severe) may correlate with Laser recommendations, making them easier to identify.

The RAG-based reward shaping likely improved learning speed by providing extra guidance. By rewarding actions consistent with similar past cases, we injected domain knowledge into the reward. Reward shaping is known to speed up RL training; although we did not separately quantify its effect here, the shaped reward prevented extremely sparse learning. In a real deployment, one could extend this by retrieving from external medical knowledge bases or guidelines to validate decisions.

An AI agent(1,12) that recommends treatments could assist clinicians in decision-making, particularly in settings where specialists are scarce. The approach here could be integrated into a decision support system: a new patient's features are fed into the (federated) model to suggest laser or conventional therapy. RL-based systems naturally adapt policies as more outcomes accumulate, potentially personalizing therapy. In periodontology (9–11), even a modestly accurate system could improve the consistency of care or highlight borderline cases for review. Performance was only moderate, reflecting the

limitations of the data. With only 300 cases total and a small test set, statistical variance is high. The model's accuracy may improve with larger, more diverse datasets. Our reward shaping was simplistic; more sophisticated RAG techniques (using knowledge graphs or LLMs) might yield better guidance. We also did not incorporate temporal outcomes (the data was cross-sectional). True treatment response is sequential; a more advanced RL could model multi-stage therapy outcome. Lasers provide some benefit but are not significantly better than SRP. The model's uncertainty between classes reflects clinical ambiguity. Our federated framework addresses privacy concerns in line with the recognized need for multi-center learning. The DQN architecture (a 32-unit MLP) is simple yet effective for tabular clinical data, similar to prior clinical RL models(13–15). In the future, the dataset will include real-world clinical records from multiple centers, with additional variables such as radiographic and microbiome features. Real-time reinforcement learning will enable adaptive decision-making and improve clarity by utilizing advanced methods, such as counterfactual analysis. Some limitations of this study include the small sample size, potential bias toward laser treatment due to the assignment of labels, and the use of predefined PCA components, which may complicate the interpretation of raw features. The federated setup based on simulation may also differ from actual network conditions.

CONCLUSION

We presented a federated Deep Q-Network recommending laser therapy versus conventional treatment based on patient features. Self-supervised PCA reduced features to 8, and RAG-inspired reward shaping provided training signals. The model achieved 60% accuracy and a ROC AUC of ~0.69. Key points include the significance of feature combinations and multi-institutional RL training without data sharing. This showcases the potential of RL, federated learning, and self-supervision in medical decision support. Future work will expand the dataset, refine reward models, and test in clinical workflows.

DECLARATIONS

Author Contributions

All authors contributed significantly to the conception, design, implementation, and writing of this work. All authors reviewed and approved the final manuscript.

Funding

Not Applicable.

Conflicts of Interest

The authors declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

This article does not contain any studies involving human participants or animals performed by any of the authors.

REFERENCES

1. AbdelAziz NM, Fouad GA, Al-Saeed S, Fawzy AM. Deep Q-Network (DQN) Model for Disease Prediction Using Electronic Health Records (EHRs). Sci [Internet]. 2025;7(1). Available from: https://www.mdpi.com/2413-4155/7/1/14

2. Jayaraman P, Desman J, Sabounchi M, Nadkarni GN, Sakhuja A. A Primer on Reinforcement Learning in Medicine for Clinicians. NPJ Digit Med [Internet]. 2024;7(1):337. Available from: https://doi.org/10.1038/s41746-024-01316-0

3. Yadalam PK, Natarajan PM, Saeed MH, Ardila CM. Variational Approaches for Drug-Disease-Gene Links in Periodontal Inflammation. Int Dent J [Internet]. 2024; Available from: https://www.sciencedirect.com/science/article/pii/S0020653924015375

4. Yadalam PK, Natarajan PM, Mosaddad SA, Heboyan A. Graph neural networks-based prediction of drug gene association of P2X receptors in periodontal pain. J Oral Biol Craniofac Res. 2024;14(3):335–8.

5. Yadalam PK, Arumuganainar D, Ronsivalle V, Di Blasio M, Badnjevic A, Marrapodi MM, et al. Prediction of interactomic hub genes in PBMC cells in type 2 diabetes mellitus, dyslipidemia, and periodontitis. BMC Oral Health. 2024 Mar;24(1):385.

6. Yadalam Raghavendra V.; Ramadoss, Ramya; Shrivastava, Deepti; Alruwaili, Awsaf Murdhi; Faheemuddin, Muhammad; Srivastava, Kumar Chandan PK; A. Identification of Repurposed FDA Drugs by Targeting Sclerostin via the Wnt Pathway for Alveolar Bone Formation. European J Gen Dent [Internet]. 2024;(EFirst). Available from:http://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0043-1777841

7. Yang J, El-Bouri R, O'Donoghue O, Lachapelle AS, Soltan AAS, Eyre DW, et al. Deep reinforcement learning for multi-class imbalanced training: applications in healthcare. Mach Learn. 2024;113(5):2655–74.

8. Liu M, Wang S, Chen H, Liu Y. A pilot study of a deep learning approach to detect marginal bone loss around implants. BMC Oral Health. 2022 Jan;22(1):11.

9. Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. J Med Internet Res. 2020 Jul;22(7):e18477.

10. Baker QB, Swedat S, Aleesa K. Automatic Disease Diagnosis System Using Deep Q-Network Reinforcement Learning. In: 2023 14th International Conference on Information and Communication Systems (ICICS). 2023. p. 1–6.

11. Kim Y, Suescun J, Schiess MC, Jiang X. Computational medication regimen for Parkinson's disease using reinforcement learning. Sci Rep. 2021 Apr;11(1):9313.

12. Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Stat Med. 2009 Nov;28(26):3294–315.

13. Gao X, Li D, Duan Y, Wu L. Risk factors and prediction model of peri-implantitis in post operative periodontitis patients. Am J Transl Res. 2024;16(9):4741–50.

14. Rekawek P, Herbst EA, Suri A, Ford BP, Rajapakse CS, Panchal N. Machine Learning and Artificial Intelligence: A Web-Based Implant Failure and Peri-implantitis Prediction Model for Clinicians. Int J Oral Maxillofac Implants. 2023;38(3):576–582b.

15. Lee WF, Day MY, Fang CY, Nataraj V, Wen SC, Chang WJ, et al. Establishing a novel deep learning model for detecting peri-implantitis. J Dent Sci. 2024 Apr;19(2):1165–73